NTP
National Toxicology Program

Analysis of NTP's tumor incidence data from 2-year cancer bioassay: New Methods

Shyamal D. Peddada, Ph.D.
National Institute of Environmental Health Sciences

NTP Board of Scientific Counselors
April 13, 2011

Role of Statistical Methodology

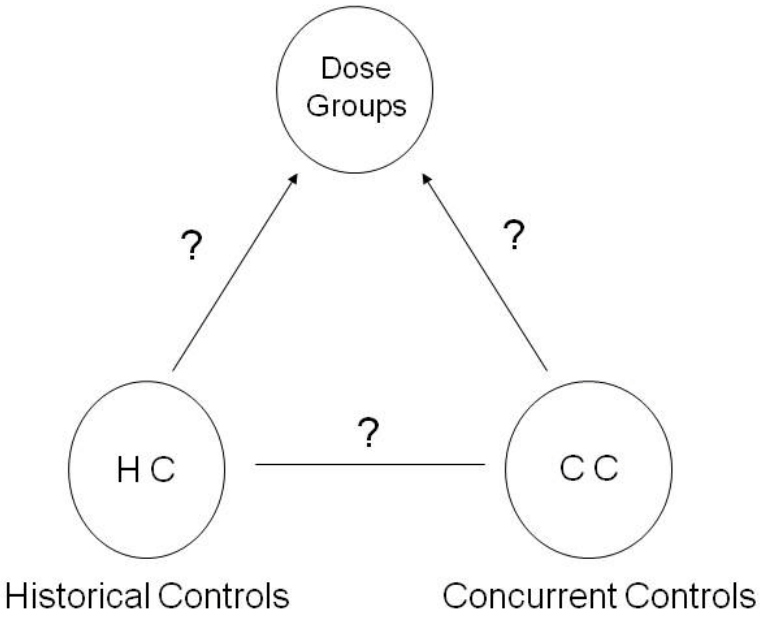
- The NTP uses a variety of information, including the p-values derived from statistical procedures, to make a decision regarding carcinogenicity

- Thus, the p-values are not the ***only*** piece of information used when making decisions

Outline

- The paradigm and desired comparisons
- Max-Iso-Poly-3 test
  - Proposed statistical test
  - Examples
- Use of historical control data
  - Proposed statistical tests
  - Examples

The paradigm



Desired comparisons

- Dose groups vs. CC
  - Trend
  - Pairwise comparisons of individual dose groups vs. CC
- Dose groups vs. HC
  - Trend
  - Pairwise comparisons of individual dose groups vs. HC
- CC vs. HC

Comparison of dose groups vs. concurrent control …

Methodology presently used…

The Poly-3 trend test

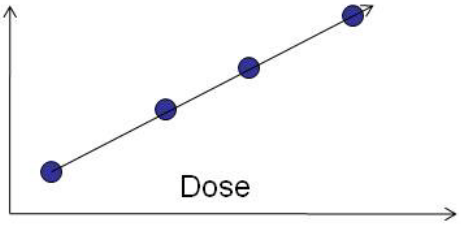
- Based on the well-known Cochran-Armitage trend test for binary data
- Survival differences among dose groups is accounted by using Poly-3 correction to sample size
  - Bailer & Portier (1988)

Poly-3 correction to sample size

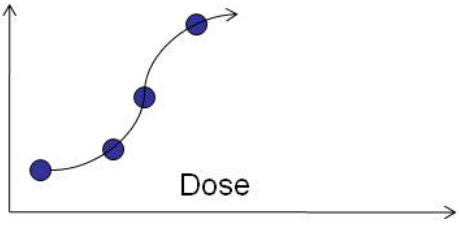
- An animal contributes 1 to the sample size if either
  - It survived until the end of the study
  - Or it died with the tumor in question
- Otherwise it contributes $\left(\frac{d}{T}\right)^3$ to the sample size
  - d is the duration of time the animal survived
  - T is the length of the study (typically 2 years).

Why propose an alternative to the Poly-3 trend test?

- Intrinsically the Poly-3 trend test is ideal when the trend is linear in dose



- Often loses power for non-linear monotonic trends



Some underlying principles

Develop a new trend test such that:

- Controls false positive rate as well as the Poly-3 trend test
- As powerful as Poly-3 trend test for linear trends
- Has greater power than Poly-3 trend test for non-linear monotonic trends
- No complicated modeling or other assumptions are made

Max-Iso-Poly-3: The proposed test

Two components

- T1: Isotonic regression based test (Williams' type test)
- T2: Poly-3 trend test

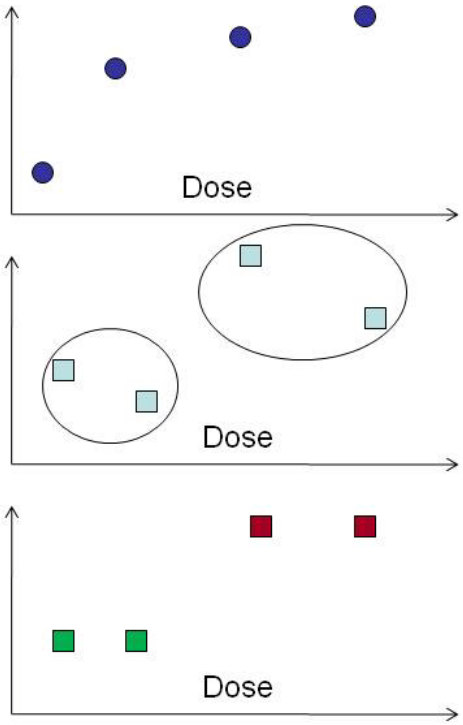
Max-Iso-Poly-3 test statistic
= Max (T1, T2)

Some features of T1

- Modeling assumptions:
  - No complicated model used to describe dose-response relationship
  - Uses mathematical inequalities to describe monotonicity
- Survival differences among dose groups:
  - Like NTP's Poly-3 trend test, it uses Poly-3 corrections to sample size
  - No additional assumptions are made
- Computationally simple

Isotonic regression

- Hypothesized trend (monotonic trend in dose)
- Observed data
- Isotonized data

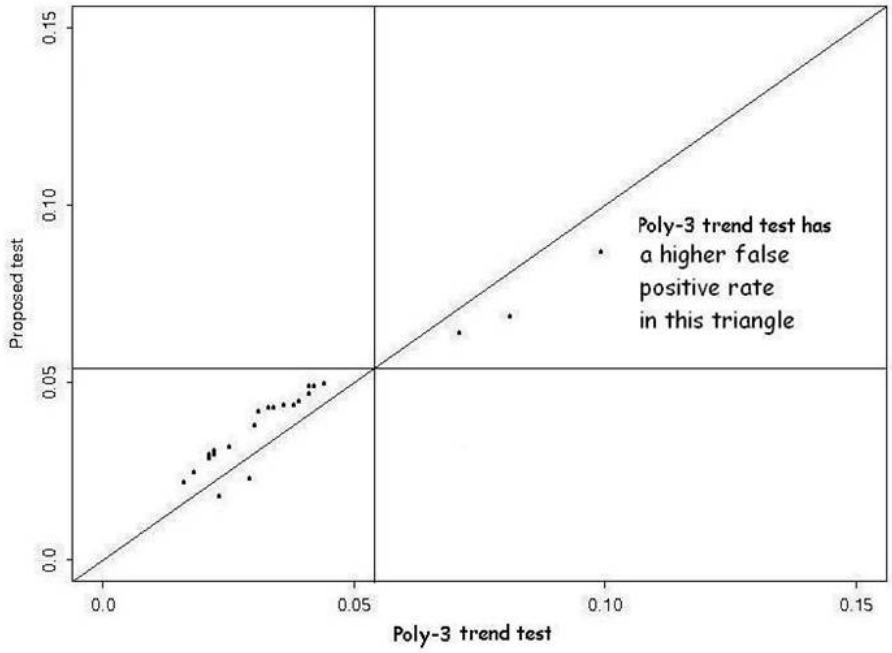


Max-Iso-Poly-3 vs. NTP's Poly-3 trend test

Simulation studies…

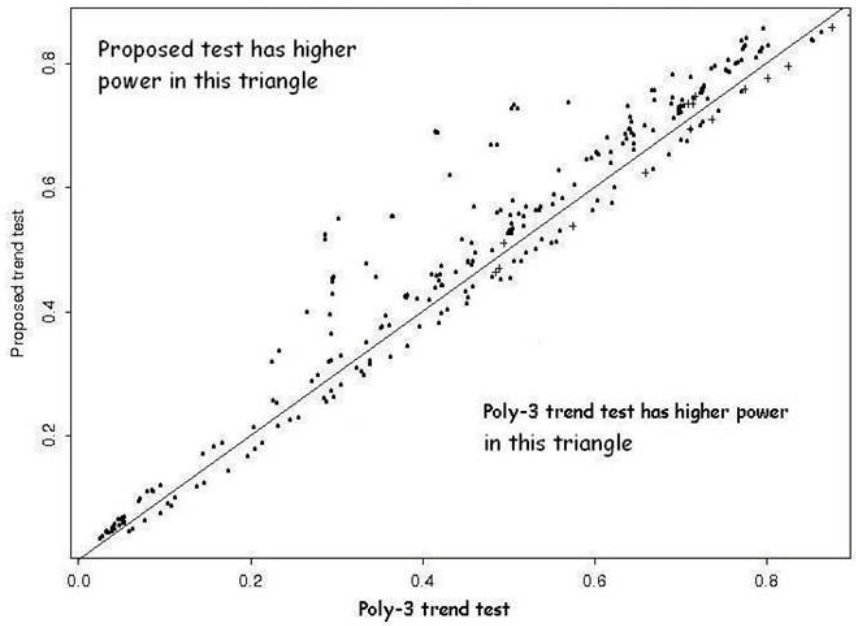
False Positive Rates

Figure 1



Power Comparisons

Figure 2



Summary

Max-Iso-Poly-3 test relative to Poly-3 test

- Controls the false positive rate as well as the Poly-3 trend test
- As powerful for linear trends
  - At most a 10% loss in power (e.g. 45.4% vs. 50.2 %)
- Increased power for non-linear monotonic trends
  - As much as 66% gain in power in comparison to NTP's Poly-3 test (e.g. 69% vs. 41.5%)

Pairwise comparisons of individual dose groups vs. CC

Compare each dose group with the control group using the NTP's current strategy but replace the Poly-3 test for 2 groups by Max-Iso-Poly-3 test for 2 groups

An Example…

Chemical: Isoprene

Mammary Gland Fibroadenoma: Female Rats

- Tumor rates: (19/50, 35/50, 32/50, 32/50)
- Poly-3 percents: (42.9, 74.3, 73.7, 73.2)

Trend tests:

- Poly-3 trend test: P-value = 0.11
- Max-Iso-Poly-3 P-value < 0.001

Pairwise comparisons with control:

Poly-3: (---, <0.001, < 0.01, < 0.01)

Max-Iso-Poly-3: (---, <0.001, <0.001, <0.001)

Concluding remarks on Max-Iso-Poly-3 trend test

- Controls the false positive rate as well as the Poly-3 trend test
- As powerful as Poly-3 trend test for linear trends
- Substantial gains in power over Poly-3 trend test for non-linear monotonic trends
- No complicated modeling or assumptions other than what is assumed by the Poly-3 trend test

Using historical control data …

Two possible strategies

- Strategy 1: One single global comparison:

  Compare dose groups with all controls (CC and HC) together, while acknowledging variability among HC

- Strategy 2: Make three separate comparisons:

  1. Compare dose groups with CC
  2. Compare dose groups with HC
  3. Compare CC with HC

Comparison of the two strategies

- Strategy 1
  - Provides a single p-value that takes into account CC and HC
  - Consistent with earlier attempts by statisticians
- Strategy 2
  - Three pieces of information which can be used along with all other information in making decisions
  - NTP's preferred strategy

Current NTP strategy

- Formally compare dose groups with CC
  - Poly-3 test
- Informally compare dose groups with HC
  - Use historical control range
- Informally compare CC with HC
  - Use historical control range

Comments on historical control range

- No clearly established methodology
- Requires a "large" number of studies
  - Hard to implement when a new strain of rodents is used
- Information on intermediate values is ignored, only the extreme data are used
  - A single outlier can make the range very large
- High false positive rate when comparing CC vs. HC
  - With 5 HC groups, false positive rate = 33%
  - Would need 39 HC groups for false positive rate of 5%

Proposed strategy

- Compare dose groups with CC
  - Max-Iso-Poly-3 test
- Compare dose groups with HC
  - Max-Iso-Poly-3 type test that accounts for within and between group variability of historical controls
    - Based on Peddada et al. (2007, JASA)
- Compare CC with HC
  - Z-test that accounts for within and between group variability of historical controls

Important Assumption

- All controls are assumed to be from a common homogenous population, although there may be variability among the controls

Homogeneity among controls (Standard practice by NTP)

All controls are matched in terms of various characteristics, such as:

- Sex & species & strain
- Tumor & tissue type
- Route of exposure (although sometimes NTP also uses all routes combined)
- Date of study (Five year sliding window)
- Diet
- Etc

Pairwise comparisons of individual dose groups vs. HC

Compare each dose group with the HC using the NTP's current strategy but using Max-Iso-Poly-3 type test for 2 groups

Two Examples ...

Tetralin:
Male Rats, Kidney - Renal Tubule Adenoma

- Tumor rates: (0/50,1/50, 1/50, 2/50)
- Poly-3 percents: (0, 2.3, 2.4, 4.6)
- Historical control incidence (6 studies): 2/299

Tetralin:
Male Rats, Kidney - Renal Tubule Adenoma

Comparison	P-value
CC vs. Doses	0.17 (Poly 3)
	0.15 (Max-Iso-Poly 3)
HC vs. Doses	0.017
CC vs. HC	0.58

No significant difference between HC and CC

Dose groups are significantly different from HC (P = 0.017)

This shows that the HC data can be useful in interpreting data on rare tumors.

Androstenedione: Female Rats, Mononuclear Cell Leukemia

- Tumor rates: (5/50, 11/50, 18/50, 15/50)
- Poly-3 percents: (10.4, 23.3, 38.4, 30.3)
- Historical control incidence (Only one study): 12/50 (24%)

Androstenedione: Female Rats, Mononuclear Cell Leukemia

Comparison	P-value
CC vs. Dose	0.029 (Poly 3)
	0.005 (Max-Iso-Poly 3)
HC vs. Dose	0.25
CC vs. HC	0.05

HC and CC are marginally significantly different (P= 0.05)

Dose groups significantly different from CC (P = 0.005)

Dose groups not significantly different from HC (P = 0.25)

Thus, even though we have only one HC study, we are able to perform a valid statistical inference

Perhaps, in this case CC is more relevant than HC

Concluding remarks on the proposed HC tests

- No complicated models are used
- No more assumptions are made than what is currently being assumed by the NTP
- Applicable even when the number of HC is as small as one group
- Better control of false positive rates
- Powerful for rare tumors

Questions?

Some examples of power where at least one test has a power > 80%

Example	Poly-3 trend test (Power in %)	Max-Iso-Poly-3 trend test (Power in %)	Percent gain in power relative to Poly-3 trend test
1	63	89	41
2	69	95	39
3	76	94	24
4	86	83	- 3
5	89	86	- 3.3

Some examples of power where at least one test has power between 70% and 80%

Example	Poly-3 trend test (Power in %)	Max-Iso-Poly-3 trend test (Power in %)	Percent gain in power relative to Poly-3 trend test
1	47	73	53
2	48	72	50
3	57	75	32
4	77	74	- 4
5	71	68	- 4.5

Top 3 spontaneous neoplasms out of >250 types

Male mice Tumor (%)	Female mice Tumor (%)	Male rats Tumor (%)	Female rats Tumor (%)
Hepatocell. Adenoma (55)	Hepatocell. Adenoma (27)	Testes Interstitial Cell Adenoma (81)	Pituitary gland: Pars distalis Adenoma (55)
Hepatocell. Carcinoma (31)	Malignant Lymphoma (20)	Pituitary gland: Pars distalis Adenoma (47)	Mammary gland Fibroadenoma (52)
Lung A/B Adenoma (15)	Hepatocell. Carcinoma (10)	Leukemia (39)	Leukemia (21)